

LAB 02

Wordcount program in Hadoop using Cloudera platform.

In this practical Lab 2 session, you will complete two activities.

Activity 1

Kindly follow the link below to run your first wordcount program:<https://www.youtube.com/watch?v=Wb5p8S5jZCc&feature=youtu.be>

Please follow each step in the video to understand how the wordcount program works.

Do not make any changes at this point in your practical steps.

Provide evidence of the output screen in the team chat for completion.

Guide and Resources:

Open Eclipse inside the Cloudera platform.

Create a Java file as per the video.

Copy the source code from the following link:

https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html#Source_Code

Open the terminal and create a simple text file.

Send the file to HDFS.

Run the HDFS syntax to execute the wordcount.

Activity - 2

Refer to Spectrum for the dataset for MapReduce (.txt) and copy this file into your Cloudera text file. Run this file for wordcount.

You need to attach evidence of completion in the Spectrum Practical-2 Submission link.

Save and submit your Lab documents in PDF with the following filename format.

- a. Screenshot 1 – Text file
- b. Screenshot 2 – transferred to hdfs
- c. Screenshot 3 – Output of map reduce.

Save and submit your Lab documents in **PDF** with the following filename format.

Name_ID_Lab2	E.g.: E.g.: Riyaz-123456-Lab2.pdf
--------------	-----------------------------------